

Final Technical Report

Arizona Sample and Data Inventory

By M. Lee Allison, Principal Investigator, Stephen M. Richard, and Aric
V. Villarreal

Prepared for the USGS National Geological and Geophysical Data Pres-
ervation Program

under

USGS Award 07HQGR0124

January, 2009

TABLE OF CONTENTS

Introduction.....	3
Scope and implementation.....	3
Metadata	3
Methodology.....	4
Results.....	4
Appendix I . Definition of metadata elements.....	8
Administrative:	8
Descriptive:.....	9
Structural.....	10

Introduction

As part of the Energy Policy Act of 2005, the U.S. Geological Survey was tasked with creating a National Geological and Geophysical Data Preservation Program (NGGDPP). This Program is intended to work principally with the nation's geological surveys to: (1) archive geologic, geophysical, and engineering data, maps, well logs, and samples; (2) provide a national catalog of such archival material; and (3) provide technical and financial assistance related to the archival material.

The Program is envisioned as a national network of cooperating geoscience materials and data repositories that are operated independently yet guided by common standards, procedures, and protocols for metadata. The holdings of all collections will be widely accessible through a single, common, and mirrored Internet-based catalog, the National Digital Catalog, thus maximizing the availability of and interconnectedness of all the collections.

Scope and implementation

In 2007, the USGS provided \$5,000 in matching funds to the Arizona Geological Survey to conduct an inventory of AZGS geological and geophysical data and sample holdings in its Tucson offices for the NGGDPP. These holdings include print and digital data, as well as physical artifacts. The inventory was collected from March to June 2008, and entered in Microsoft Excel spreadsheets, which were subsequently consolidated into a Microsoft Access database. The inventory categories and characterizations were derived from the NGGDPP data entry template for populating the National Digital Catalog. Inventoried collections were then analyzed to determine how to map the data into the online collection inventory.

This inventory comprises Phase I of the AZGS involvement in the NGGDPP, in which all extant holdings at AZGS have been described on the *collection* level. This means that, rather than creating metadata for every single object, as occurs for a typical library catalogue, metadata is generated describing whole groups of objects. The purpose of this approach is twofold: first, to determine generally the type and quantity of all data held by AZGS; and second, to determine which objects or collections will be singled out for more in-depth descriptive or preservation projects in the future.

Metadata

Metadata is "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."¹ Any resource, for instance, a map, can be counted as data. Metadata content describes the particular artifacts (digital or physical) so that they can be found, retrieved, and cited. Metadata for a map could include its date of publication, author, publisher, scale, size, media (whether it is a digital or print object, for example), geographic area covered, edition, longitude and latitude coordinates, title, type (geologic, topographic, political, etc.), and other terms.

¹ NISO, *Understanding Metadata*. (Bethesda, MD: NISO Press, 2004).
<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

Metadata content is structured into a metadata schema, which provides a uniform template for facilitating the retrieval, management, and preservation of resources, regardless of format.

Methodology

The inventory was conducted by describing *collections* of artifacts in the possession of the Arizona Geological Survey. Collections have been initially defined and categorized according to either location within AZGS premises, or by way of some attribute intrinsic to a group of resources that unifies them in a collection. For example, aerial photographs collected together in a steel case drawer would be counted, and then assigned a collection title corresponding to the label on the drawer. If that label is too generic to lend itself to an accurate description of the items under purview, then the collection title would correspond to attributes drawn from the data sources themselves. The collection in the file drawer marked “1955 Photos” was given the title “1955 aerial photos—Tucson, Nogales, Phoenix, Mesa” to better describe the collection in terms of 1) media, 2) date of publication, and 3) geographic area covered. “1955 Photos” was placed in the “References” field at the end of the Excel spreadsheet entry for this particular collection, as well as the particular physical location for that collection. The entire “References” entry would therefore read “See metal file case in GIS room, with drawer marked ‘1955 photos,’” as a complete structural metadata entry. Inventoried resources are all located at the AZGS headquarters at 416 West Congress St, Tucson Arizona, except for rock core, which are held in Room B-005 in the state building next door at 400 West Congress St.

These initial collections were redefined to aggregate related kinds of resources that might be physically located in different parts of our building into single collections. Collection type key words, and time period of content information were added, and the compilation spreadsheets were aggregated into a single Microsoft Access database.

Counting units were defined for each collection type, and an estimates of the number of objects in each collection was made. Core and cutting samples were counted primarily by the number of wells represented, with the sampled interval length (in linear feet) in parentheses following the number of wells in the “Quant?” element. The number of wells for both Oil and Gas and water well cuttings is given, but there is no estimation of sample quantity, since to physically count the number of cutting samples is beyond the scope of this project. Other exceptions include some large aerial photograph and microform collections, in which the objects were collected into envelopes. In order to attain a more expeditious, as well as more structurally descriptive count of the objects within these collections, only the envelopes were counted, with the average number of photographs per envelope given in the “Unit of Measurement” element. All of the objects to be found in the AZGS library catalogue were counted as one collection.

Results

Table 1 summarizes 24 collections registered in the national catalog using the data input web interface. Ongoing cataloging and inventory work at AzGS under the auspices of phase 2 funding are resulting in definition of some more specific sub collections within these initially identified collections, and we will continue to evolve our cataloging scheme as the project continues. This initial inventory work will provide critical input

for development of our long term data preservation plan during the current funding cycle.

Table 1. AzGS collections registered with National Catalog

ID	Collection-Type	Collection-Category	Name	Description
P1418	Derived and Indirect Geo-science data	Drilling/completion reports	Drilling/completion reports	14 file folders of drilling/completion reports.
P1419	Derived and Indirect Geo-science data	Field notes	Field notes	Jon Spencer, Stephen Richard field notes from Geologic mapping 1980-present, In file cabinets, about 6 drawers . Miscellaneous other field notebooks from other AZGS geologists, quantity hard to assess.
P1420	Physical Geo-science	Fluid samples	Fluid samples	7 boxes of well water samples. Not clear how well documented collection location or sampling procedure is.
P1421	Derived and Indirect Geo-science data	Geochemical Data	Geochemical data	Damon Geochron data, Radon spectrometer data (don't include if in AZGS pub), TOC reports filed by Standard Oil and Arco.
P1422	Derived and Indirect Geo-science data	Maps	Maps I	map drawers with maps compiled for geothermal energy assessment, 10 drawers of 'AzGS map files'. Needs to be inventoried in more detail to determine how much of this is worth cataloging for public discovery.
P1423	Derived and Indirect Geo-science data	Maps	Maps II	2 boxes of maps, scope and content uncertain
P1424	Derived and Indirect Geo-science data	Maps	Maps, geologic I	Santa Fe Railroad geologic maps of railroad lands, miscellaneous unpublished geologic maps
P1425	Derived and Indirect Geo-science data	Geophysical Data	Maps, geologic II	gamma radiation mylars, origin uncertain, other maps and data with unknown content, needs to be inventories
P1426	Derived and Indirect Geo-science data	Maps	Superconducting Supercollider site assessment data	Maps, other ancillary data collected during assessment of Maricopa Mountains and Sierra Mountains proposed sites for Superconducting Supercollider
P1427	Derived and Indirect Geo-science data	Maps	Uranium Files	unpublished maps compiled for uranium resource assessment, 1980
P1428	Derived and Indirect Geo-science data	Special analysis data	Aggregate Materials inventory sites	Maps showing sampled materials pits, tables of aggregate quality data compiled by ADOT

P1429	Derived and Indirect Geo-science data	Paper reports	Mining history	files on individual mines, mining company reports, newspaper clippings reporting on mining activity
P1430	Derived and Indirect Geo-science data	Paper reports	Oil and Gas Files	Permits to drill, drilling records, oil and gas commission meeting minutes, liquified gas storage facility files
P1431	Derived and Indirect Geo-science data	Paper reports	Paper reports	Collections of folders of information on flooding in Arizona, files from Eldred Wilson. These are probably unrelated, but lumped here until we inventory them more closely to classify better
P1432	Derived and Indirect Geo-science data	Photographs	Air photos	Aerial photographs, as old as 1936, various photo sizes and scales, items counted are file folders or envelopes, generally containing one flight line. There are numerous other groups of photos not in envelopes, totaling about 2800 more photos.
P1433	Derived and Indirect Geo-science data	Photographs	Field Photographs	Count is a minimum. Most are in possession of geologists who took photos, most are film transparencies; geologists now take digital photos, and we didn't even try to count those.
P1434	Physical Geo-science	Rock cores	Rock cores	Skeletonized core in boxes, mostly from mineral exploration projects, and from 3 oil wells, one geothermal exploration borehole
P1435	Physical Geo-science	Rock cuttings	Rock cuttings	rock chips in vials or plastic bags. Material in vials, mostly from water wells, has been washed. Material in plastic bags, mostly from oil and gas wells, is not washed and is stored in boxes.
P1436	Derived and Indirect Geo-science data	Scout tickets	Scout tickets	3 by 5 index cards, in index card file drawers, 40 linear inches of drawer space with cards, mostly from 1950's. Organized by location.
P1437	Physical Geo-science	Thin sections & polished sections	Thin sections & polished sections	Glass thin sections, a few polished sections. Mostly from samples collected for lithologic characterization in connection with mapping projects.
P1438	Derived and Indirect Geo-science data	Lithology logs	Lithology logs, Mineral exploration projects	Star Hill material in binders with photos of material as well as standard lithology logs, Courtland-Gleeson data includes some assay data, Red Mountain data from extensive exploration of porphyry copper system there.
P1439	Derived and Indirect Geo-science data	Lithology logs	Lithology logs, basement files	file folders for boreholes, mostly water wells. Information in folders is variable, some driller's logs, some e-logs, some just a registration form. Not sure what we've got here.

P1440	Derived and Indirect Geo-science data	Well logs	Oil and Gas bore-hole files	Each well has a file containing logs from the well, as well as the permit information and completion reports
P1441	Derived and Indirect Geo-science data	Well logs	Oil and Gas bore-hole logs, digital	All material associated with Oil and Gas wells has been scanned, and is in electronic form

Acknowledgements. AZGS geologists Steve Rauzi and Mike Conway provided substantial help in identifying collections and making recommendations on inventory criteria. AZGS geologist Pam Wilkinson did preliminary assessment of data, materials and samples in state, federal, and private repositories with an emphasis on identifying items at risk of being destroyed or discarded. We thank Tamara Dickinson, head of the USGS NGGDDPP, for guidance in carrying out the project. This work was funded in part by the U.S. Geological Survey, National Geological and Geophysical Data Preservation Program, grant 07HQGR0124.

Appendix I . Definition of metadata elements

The following definitions are based on the data entry attributes for the National Digital Catalog. There are three metadata types, Administrative, Descriptive, and Structural. Under each metadata type is a listing of corresponding elements specific to the NGGDPP spreadsheets:

Administrative:

Metadata of this type is concerned with the actions taken to acquire, maintain, preserve, duplicate, and disseminate a resource. Examples include accession date, provenance, time of data entry.

Available on web?

Whether any of the contents of a collection are available online.

Condition surveyed

This element entry requires the physical condition or arrangement of the collection, with regard to access.

Current staffing (FTE's)

The type of employee who regularly accesses or maintains a particular collection. Allowable entries include: Full Time –“FT,” Part Time – “PT,” Volunteer – “V,” Contractor – “C,” Student – “S.” If no person in particular is in charge of a distinct collection, then this field is left blank.

Donated from federal agencies?

Answered as a “yes” or “no,” this element indicates whether the resource was provided by a federal agency.

Expected long term usage trend (#, type)

The frequency and kind of usage to which objects within a particular collection will be put. If no exact data is available, these estimates will be qualified as “light,” “frequent,” “heavy,” etc.

How accessed

This element entails the general location in which the particular collection being described can be found.

Improvements needed

Entered here are recommendations for preservation, improved access.

Major uses*

This element describes the purposes for which a particular collection is intended, using prescribed qualifiers. These include: “Research,” “Reference,” “Land Mgmt,” “Min-

eral” “Exploration,” “Engineering,” “Teaching,” “Hazard Mitigation,” “Oil & Gas Exploration.”

Outside users**

Entered in this element are predetermined qualifiers relating to the type of user who is not an AZGS employee for whom the objects within the collection being described are intended. These qualifiers include: “K-12,” “Regulatory Agencies,” “General Public,” “Private Sector,” “Universities,” “Other Govt. Agencies,” “Professional Researchers,” “Others.”

References

Additional information: description of exact physical location, file path, individual in charge of data, publications to which the collection refers, etc.

Restrictions?***

Who possesses clearance to access the information contained within a collection. This element is entered using these qualifiers: “Public sector,” “Private sector,” “Internal users only.”

Times/yr

The frequency of use or access for the particular collection being described, or the times of year in which it would be accessed.

Descriptive:

Metadata used to characterize an object according to its intrinsic properties. Common examples of descriptive metadata include date of publication, author, title, and format.

Affiliations—agency/org

Agencies or organizations that have contributed to the authorship or publication of any objects within a particular collection. If the resource was generated by AZGS, “n/a” would be entered.

Category

“Category” is the general title description assigned to a collection, for the purposes of the NGGDPP.

Current media

This element includes all formats, electronic, print, or physical to be counted within a particular collection.

Geographic scope

Area covered within a particular collection. If anywhere in Arizona, this element is entered as “Arizona.” For other states, postal codes are entered (“OR” for Oregon, etc.).

For other countries, abbreviations are entered. For example, Canada will be entered as "CAN," Mexico will be entered as "MEX," etc.

Overall condition

An overall assessment of the physical condition, expressed with specific qualifiers ("poor," "good," etc.).

Quant?

The number of objects included in the "Unit of measurement" element the collection contains.

Unit of measurement

Unit of measurement refers to what objects are to be counted or quantified as constitutive of a particular collection.

Structural

This type of metadata describes the relation between objects within a collection, and the relationship between the collection and the repository. Examples include physical location, and the place of an object within a series ("number 427 out of 1000 microfiche rolls," for example).

Amount

The number of receptacles (digital or physical file folders, for instance) containing the objects pertaining to a collection.

Data base URL

The URL of any online database in which constituent objects of a collection can be found. Also, if objects within a collection are listed on a website, the URL is entered here.

Increasing?

This metadata field is entered on a "yes/no" basis, depending on whether or not new objects are added to the collection.

% Catalogued

The percentage of objects within a particular collection having been described or listed in a preexisting catalogue or index.

% Documented (metadata)

This element describes the percentage of objects within a collection for which meta-data exists, whether on the object itself, or as a surrogate (i.e. within a catalogue).

% Accessible in E-database

The percentage of a collection whose metadata can be accessed in a preexisting database or spreadsheet.

Predictable?

Also entered as a “yes/no,” this concerns the regularity with which new objects are added. If “Increasing?” is entered as a “no,” then this field is entered simply as “n/a.”

Web access standards*****

How data from the collection is made interoperable, entered using these following qualifiers: “web page client,” “Z39.50 web service,” “OGC catalogue.” Others may be entered.

Bibliography and links

NISO, *Understanding Metadata*. Bethesda, MD: NISO Press, 2004.

<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

National Geological and Geophysical Data Preservation Program,

<http://datapreservation.usgs.gov>